Error (E) Covariance Diagnostics

in Ensemble Data Assimilation

Kayo Ide, Daryl Kleist, Adam Kubaryk
University of Maryland

7th EnKF Data Assimilation Workshop

May 2016

Outline

Part I. Background: E-diagnostics

- E-Vectors
- E-Dimension

Part II. Application of E-diagnostics to GFS hybrid ensemble

- Ensembles of 'operational' EnKF
- Case study

Ensemble diagnostics references (not limited to)

- Patil et al, 2001
- Ockzowski et al, 2005

(low) dimensionality and instability

- Bishop and Hodyss 2011
- Enomoto et al 2015:
- Yang et al 2015:

instability vectors

Error Characteristics by **P**

- The special ellipsoid $(\mathbf{x}-\bar{\mathbf{x}})^T \mathbf{P}^{-1} (\mathbf{x}-\bar{\mathbf{x}})=1$ can be characterized by

  $p_G(\mathbf{x})$=constant

  - $\mathbf{u}_{(n)}$: unit orthogonal vector along n-th principle axis for n=1,..,≤N
  - $\sigma_{(n)}$: corresponding std

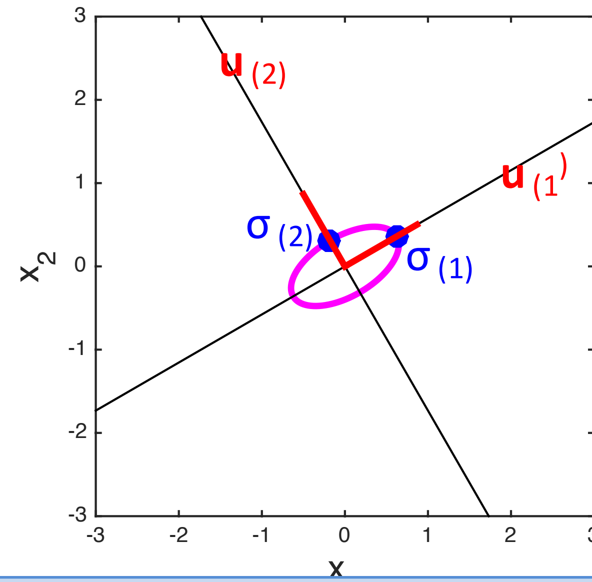
    $\sigma_{(1)} \geq \sigma_{(2)} \geq \ldots \geq \sigma_{(N)}$ [≥ 0]

    more uncertain → *less uncertain*

using SVD on $\mathbf{P}=\mathbf{U}\,\mathbf{S}^2\,\mathbf{U}^T$

$\mathbf{U}=(\mathbf{u}_{(1)},\ldots,\mathbf{u}_{(N)})$

$\mathbf{S}=\text{diag}\{(\sigma_{(1)},\ldots,\sigma_{(N)})\}$



Shape of **P** determines the uncertainty

→ The lower the mode (n small),

the larger the uncertainty ($\sigma_{(n)}$) in the corresponding direction ($\mathbf{u}_{(n)}$)

E-Diagnostics: E-Vectors on $\mathbf{X}_{en}$

- $\mathbf{X}_{en}/\sqrt{(M-1)}$ provides the same (and more) information as $\mathbf{P}_{en}$

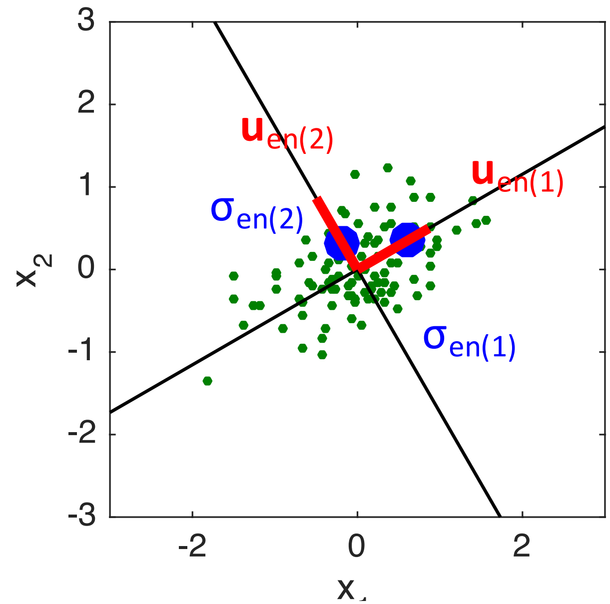
  - $\mathbf{u}_{en(n)}$: unit orthogonal vector along n-th principle axis for $n=1,..,\leq M\text{-}1$

  - $\sigma_{en(n)}$: corresponding std

$$\sigma_{en(1)} \geq \sigma_{en(2)} \geq \ldots \geq \sigma_{en(N)} \quad [\geq 0]$$

more uncertain → *less uncertain*

In the space covered by $\mathbf{U}_{en}$



using SVD on $\mathbf{X}_{en}/\sqrt{(M-1)} = \mathbf{U}_{en}\,\mathbf{S}_{en}\,\mathbf{F}_{en}^{T}$

$\mathbf{U}_{en}=(\mathbf{u}_{en(1)}, \ldots, \mathbf{u}_{en(M)})$

$\mathbf{S}_{en}=\text{diag}\,\{(\sigma_{en(1)}, \ldots, \sigma_{en(M)})\}$

$\mathbf{F}_{en}=(\mathbf{f}_{en(1)}, \ldots, \mathbf{f}_{en(M)})$

- Same concept as the EOF on $\mathbf{X}_{en}=(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_M)$ with ensemble corresponding to time series [No need to compute **P**]
- $\mathbf{F}_{en}$: projection on $\mathbf{X}_{en}$ onto $\mathbf{U}_{en}$

Error Characteristics by $\mathbf{P}_{en}$

- $\mathbf{P}_{en} = \mathbf{X}_{en}\mathbf{X}_{en}^T/(M-1)$ can be characterized by (for $M \geq N+1$)

E-vectors

- $\mathbf{u}_{en(n)}$: unit orthogonal vector along n-th principle axis for $n=1,..,\leq N$

- $\sigma_{en(n)}$: corresponding std

$$\sigma_{en(1)} \geq \sigma_{en(2)} \geq \ldots \geq \sigma_{en(N)} \quad [\geq 0]$$

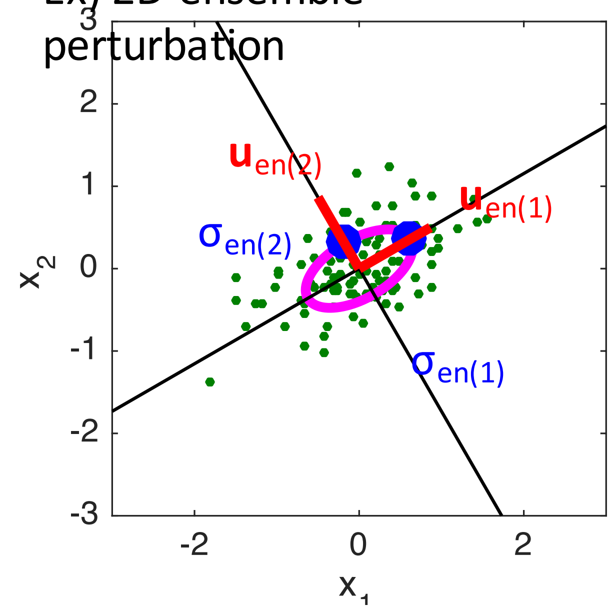
more uncertain → *less uncertain*

using SVD on $\mathbf{P}_{en} = \mathbf{U}_{en}\,\mathbf{S}_{en}^2\,\mathbf{U}_{en}^T$

$\mathbf{U}_{en} = (\mathbf{u}_{en(1)}, \ldots, \mathbf{u}_{en(M)})$

$\mathbf{S}_{en} = \mathrm{diag}\,\{(\sigma_{en(1)}, \ldots, \sigma_{en(M)})\}$

Ex) 2D ensemble perturbation



Shape of **P** determines the uncertainty

→ The lower the mode (*n* small),

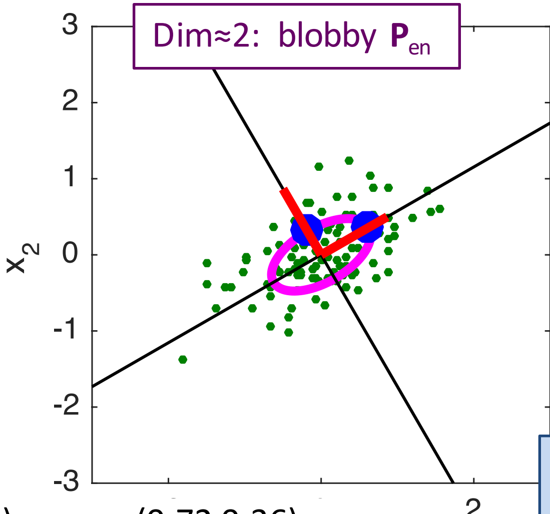
the larger the uncertainty ($\sigma_{(n)}$) in the corresponding direction ($\mathbf{u}_{(n)}$)

E-Diagnostics: E-Dimension

- E-diagnostics based on $\mathbf{P}_{en} = \mathbf{U}_{en}\ \mathbf{S}_{en}\ \mathbf{U}_{en}^{T}$ from $\mathbf{X}_{en} = \mathbf{U}_{en}\ \mathbf{S}_{en}\ \mathbf{V}_{en}^{T}$
  - Characteristics of $\mathbf{P}_{en}$
    - Shape/sizeof uncertainty
      - » $\mathbf{u}_{en(n)}$: E-vector
      - » D_{en} : E-dim
    - Size of uncertainty
      - » trace $\mathbf{P}_{en} = \sum_{n=1}^{M} (\sigma^{(n)})^2 = \sigma^2$

$$D_{en} = \frac{(\sum_{n=1}^{M-1} \sigma^{(n)} / \sigma^{(1)})^2}{\sum_{n=1}^{M-1} (\sigma^{(n)} / \sigma^{(1)})^2} = \begin{cases} M-1 & \text{if } \sigma^{(n)} = \sigma^{(1)} \quad \text{for all } n>1 \\ 1 & \text{if } \sigma^{(n)} = 0 \quad \text{for all } n>1 \end{cases}$$

Dim≈2: blobby $\mathbf{P}_{en}$



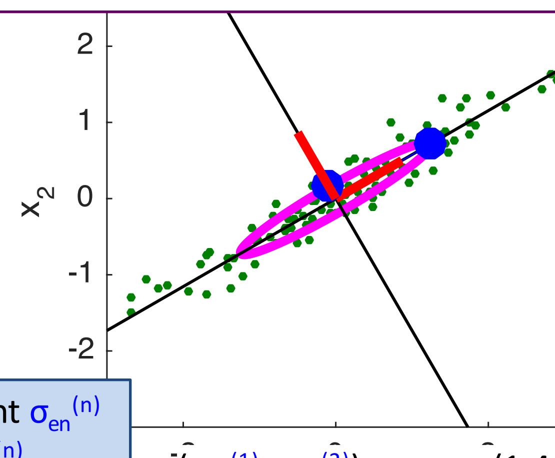
$(\sigma_{en}^{(1)}, \sigma_{en}^{(2)})$ =(0.72,0.36)

$(D_{en}$, tr $\mathbf{P}_{en}$, det $\mathbf{P}_{en})$ =(1.8,0.64,0.3)

Same $\mathbf{u}_{en}^{(n)}$ but different $\sigma_{en}^{(n)}$

→ D_{en} is defined by $\sigma_{en}^{(n)}$

Dim≈1: skinny $\mathbf{P}_{en}$ with larger uncertainty tr $\mathbf{P}_{en}$



$(\sigma_{en}^{(1)}, \sigma_{en}^{(2)})$ =(1.44,0.18)

$(D_{en}$, tr $\mathbf{P}_{en}$, det $\mathbf{P}_{en})$ = (1.24,2.1,0.3)

Data Assimilation Process

- Characteristics of practical DA process: → Lives in the space of $\mathbf{U}^b$ [$=\mathbf{U}_{en}{}^b$]
  - Analysis increment: $\Delta\mathbf{x}^a = \mathbf{K}\,\mathbf{d}$ $\quad$ $\mathbf{d} = \mathbf{y}^o - \mathbf{h}(\mathbf{x}^b)$

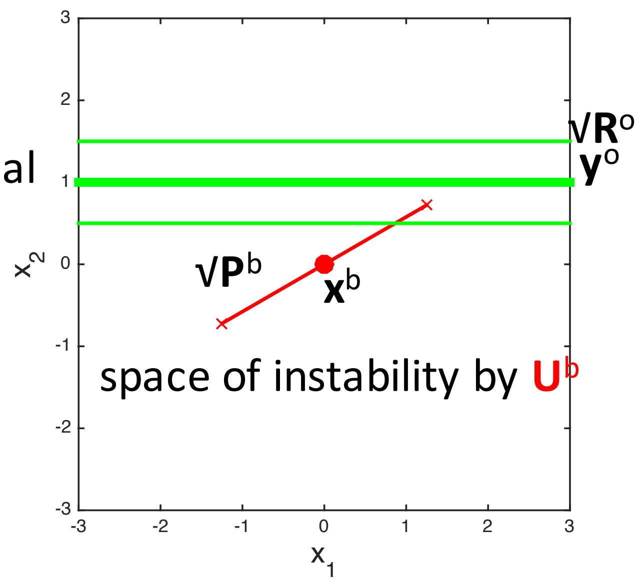
$$\Delta\mathbf{x}^a = \sum_{n=1}^{N} \alpha^a_{(n)} \mathbf{u}^b_{(n)} \qquad \mathbf{K} = \mathbf{P}^b\,\mathbf{H}^T\,(\mathbf{R}^b + \mathbf{R}^o)^{-1} \qquad \text{with } \mathbf{R}^b = \mathbf{H}\mathbf{P}^b\mathbf{H}^T$$

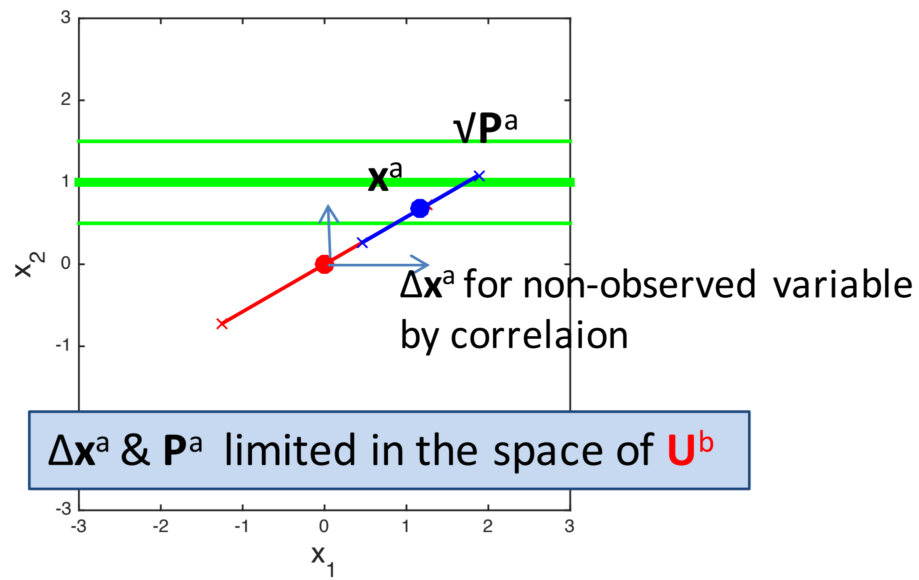
$$\alpha^a_{(n)} = (\sigma^b_{(n)})^2 (\mathbf{v}^b_{(n)})^T (\mathbf{R}^b + \mathbf{R}^o)^{-1}\mathbf{d} \qquad \mathbf{v}_{(n)} = \mathbf{H}\,\mathbf{u}_{(n)}$$

  - Analysis Error Perturbation: $\mathbf{X}^a = \mathbf{X}^b\,\mathbf{W}^a$ $\quad$: $\mathbf{W}^a = (\mathbf{I} + (\mathbf{V}^b)^T (\mathbf{R}^o)^{-1} \mathbf{V}^b)^{-1/2}$

$$\mathbf{U}^a (\mathbf{S}^a)^2 (\mathbf{U}^a)^T = \mathbf{U}^b (\mathbf{T}^a)^2 (\mathbf{U}^b)^T \qquad : (\mathbf{T}^a)^2 = \mathbf{S}^b (\mathbf{F}^b)^T (\mathbf{W}^a)^2 \mathbf{F}^b \mathbf{S}^b$$

partial obs





Data Assimilation Process

- Characteristics of practical DA process → Tends to suppress "errors of the day"
  - Analysis increment: $\Delta\mathbf{x}^a = \mathbf{K}\,\mathbf{d}$ $\quad \mathbf{d} = \mathbf{y}^o - \mathbf{h}(\mathbf{x}^b)$

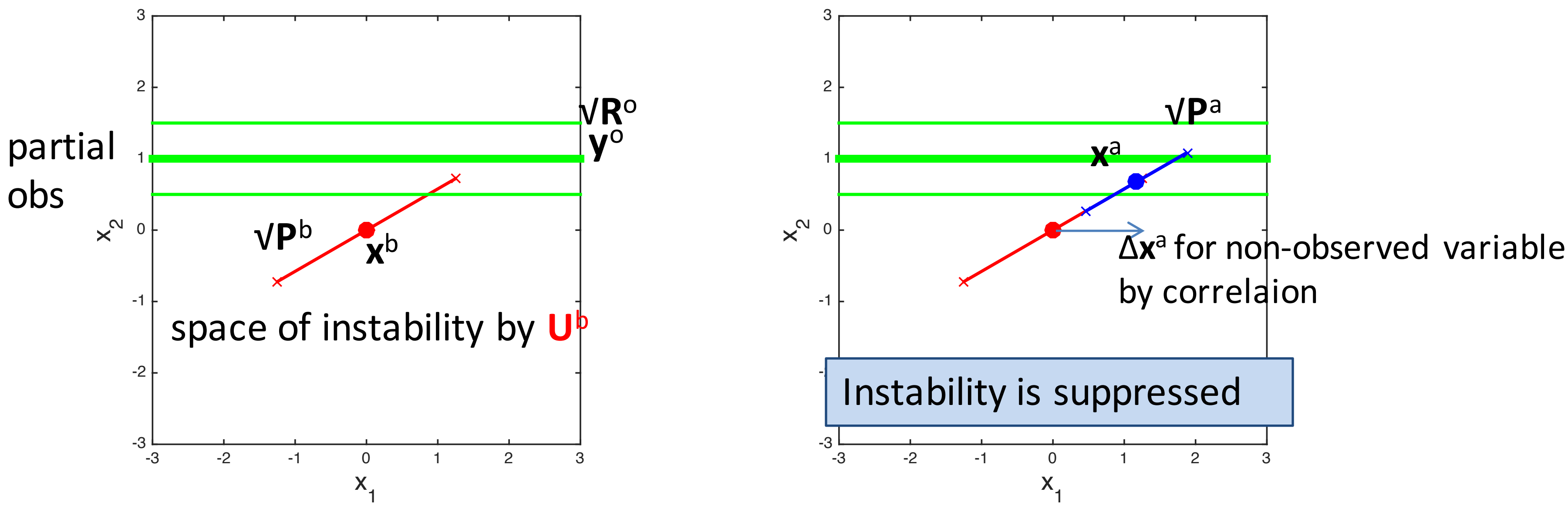
$$\Delta\mathbf{x}^a = \sum_{n=1}^{N} \alpha^a_{(n)} \mathbf{u}^b_{(n)} \qquad \mathbf{K} = \mathbf{P}^b\,\mathbf{H}^T\,(\mathbf{R}^b + \mathbf{R}^o)^{-1} \qquad \text{with } \mathbf{R}^b = \mathbf{H}\mathbf{P}^b\mathbf{H}^T$$

$$\alpha^a_{(n)} = (\sigma^b_{(n)})^2 (\mathbf{v}^b_{(n)})^T (\mathbf{R}^b + \mathbf{R}^o)^{-1}\mathbf{d} \qquad \mathbf{v}_{(n)} = \mathbf{H}\,\mathbf{u}_{(n)}$$

  - Analysis Error Perturbation: $\mathbf{X}^a = \mathbf{X}^b\,\mathbf{W}^a$ $\quad : \mathbf{W}^a = (\mathbf{I} + (\mathbf{V}^b)^T (\mathbf{R}^o)^{-1} \mathbf{V}^b)^{-1/2}$

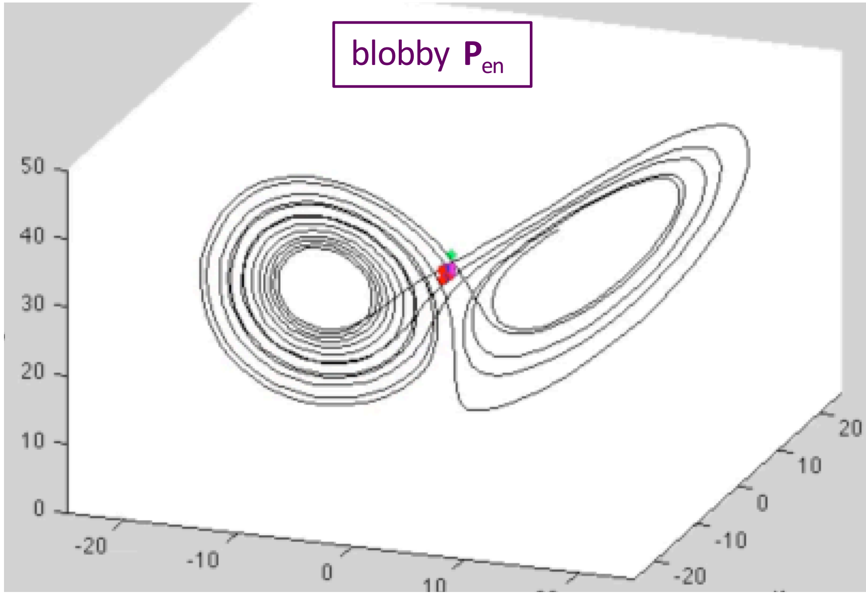
$$\mathbf{U}^a (\mathbf{S}^a)^2 (\mathbf{U}^a)^T = \mathbf{U}^b (\mathbf{T}^a)^2 (\mathbf{U}^b)^T \qquad : (\mathbf{T}^a)^2 = \mathbf{S}^b (\mathbf{F}^b)^T (\mathbf{W}^a)^2 \mathbf{F}^b \mathbf{S}^b$$

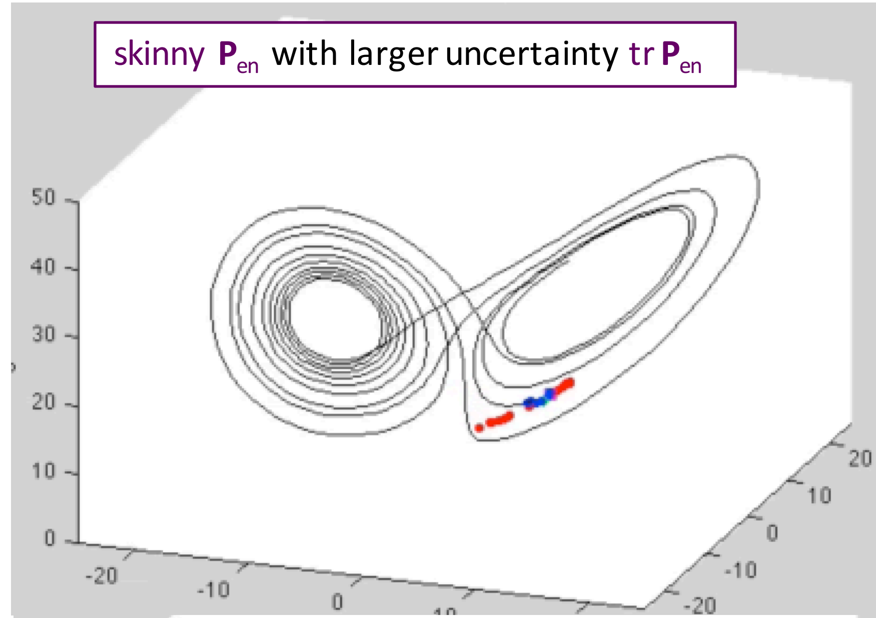
    - Tends to knock down lower order e-vectors by analysis

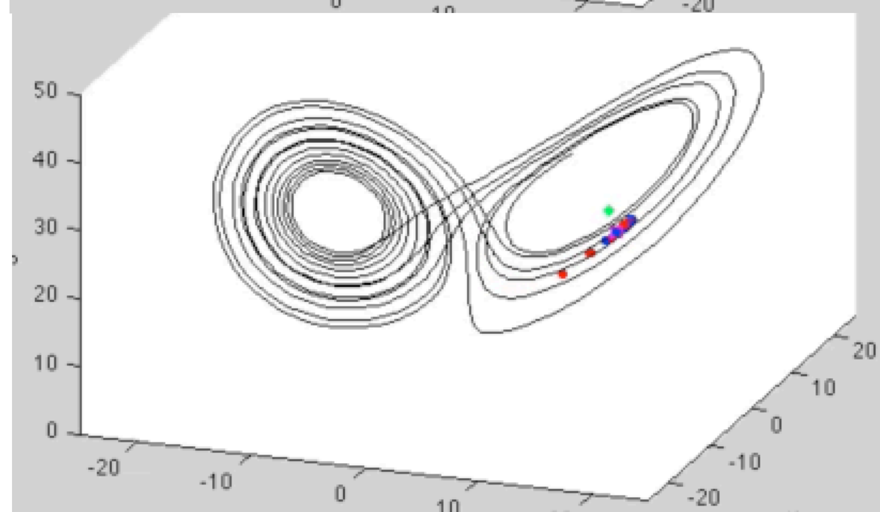


Effect of Instability on $\mathbf{P}_{en}$

- Low dimensionality ← dynamic instability in model forecast (one way)







- $\mathbf{x}_{en,m}{}^{b}$
- $\mathbf{x}_{en,m}{}^{a}$
- $\mathbf{y}^{o}$

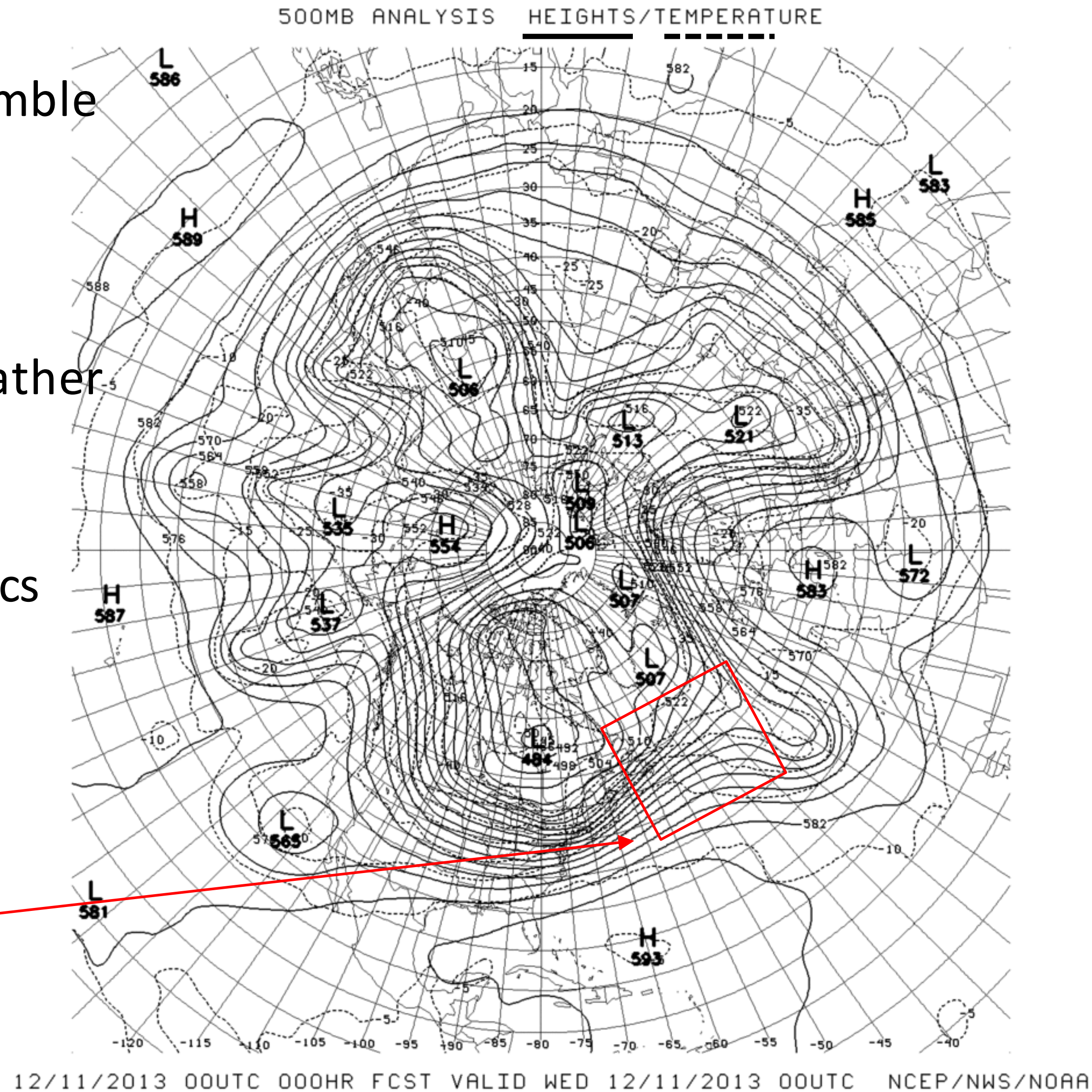
E-Diagnostics: $\mathbf{X}_{en} = \mathbf{U}_{en} \mathbf{S}_{en} \mathbf{V}_{en}^T$

- **Local E-Diagnostics**
  - For a complex system with spatio-temporal chaos, e-diagnostics should be performed locally.

- **General**
  - **E-dim based on $\mathbf{S}_{en}$** = diag $\{(\sigma_{en}^{(1)}$,…, $\sigma_{en}^{(N^*)}$=0)
    - Smaller dimension may be associated with dynamic instability and ensemble perturbation growth (reversely related to ensemble spread
  - **E-vector**: $\mathbf{U}_{en} = (\mathbf{u}_{en}^{(1)}$,…, $\mathbf{u}_{en}^{(N^*)}$) $\mathbf{u}_{en}^{(n)}$ in N-dim
    - The lower the mode, the more power it has
    - When E-dim is small, the lower mode may represent the direction of instability
  - **Projection** of $\mathbf{X}_{en}$ onto $\mathbf{U}_{en}$: $\mathbf{V}_{en} = (\mathbf{v}_{en}^{(1)}$,…, $\mathbf{v}_{en}^{(N^*)}$) $\mathbf{v}_{en}^{(n)}$ in M-dim
    - Can be used to study distribution of ensemble members

Application to Ensemble from GFS/GDAS Hybrid

- Hybrid 3DEnVar system (just replaced by Hybrid 4DEnVar)
  - T574 deterministic
  - T254 80-member ensemble
- Dates
  - 12/09 - 14/13
    [Selected for space weather Interests]
- Model state for diagnostics
  - $\mathbf{X}_{en}$: (u,v,T)
- E-diagnostics
  - Horizontal local box
    - 19x19 grid
    - 39x39 grid

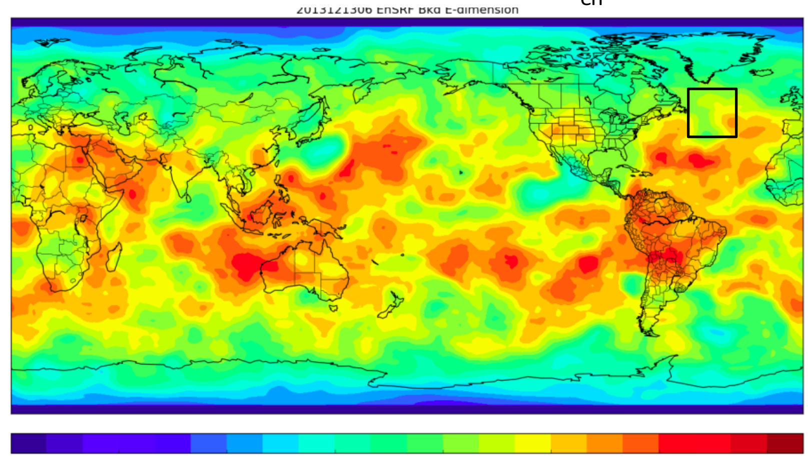


Local E-Dimension and Synoptic Conditions

- 20131209 18z: T at 500 mb
  - $\mathbf{X}_{en}$: N x M dimensional
    - N=1083=19*19*3 (u,v,T)
    - M=80 ensemble
  - Background
    - Ensemble mean
    - Ensemble spread

Low-dim dynamic instability?

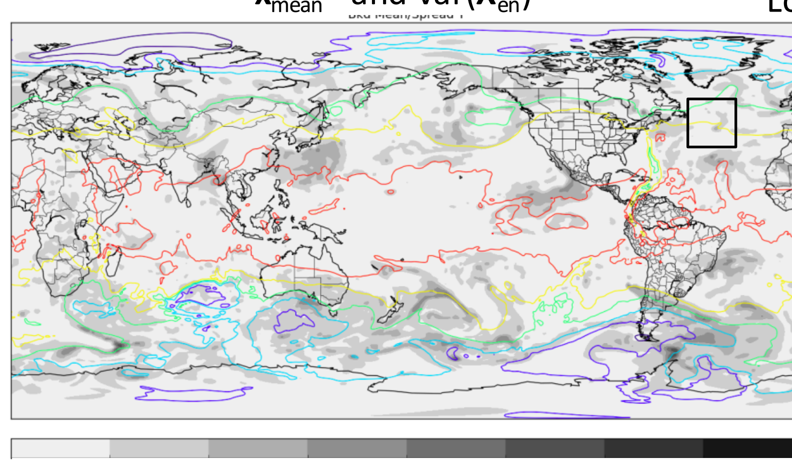
Local E-dim based on $\mathbf{X}_{en}$



Locally lower E-dim?

Locally large spread?

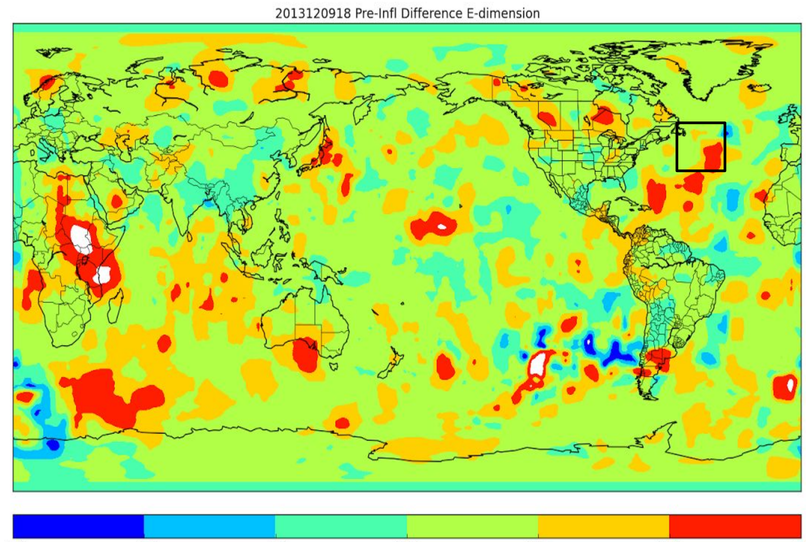
$\mathbf{x}_{mean}$ and var($\mathbf{X}_{en}$)



Large uncertainty

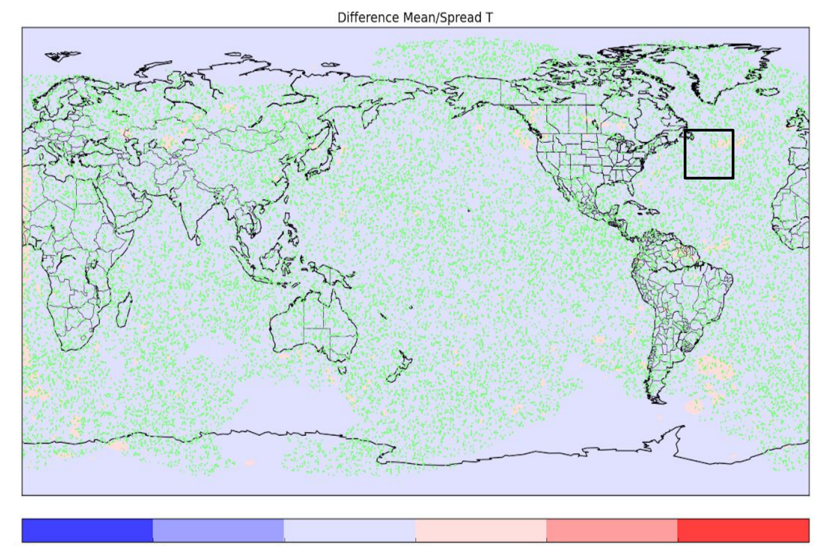
EnSRF vs LETKF: Background

- Local E-dim Difference (LETKF)- (EnSRF)



1 0 1

- Local spread std Difference

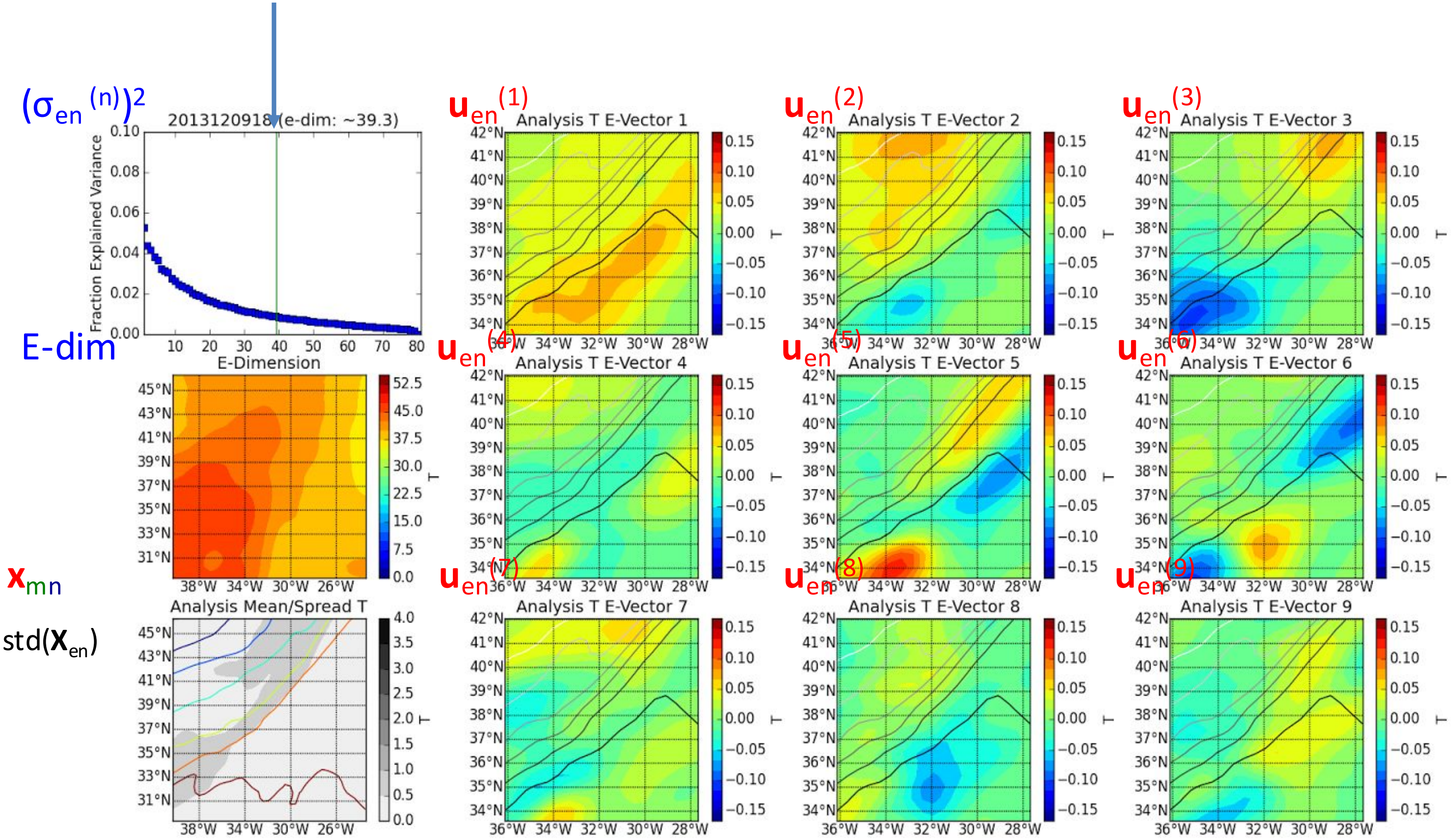


1 0 1 (K)

Very similar

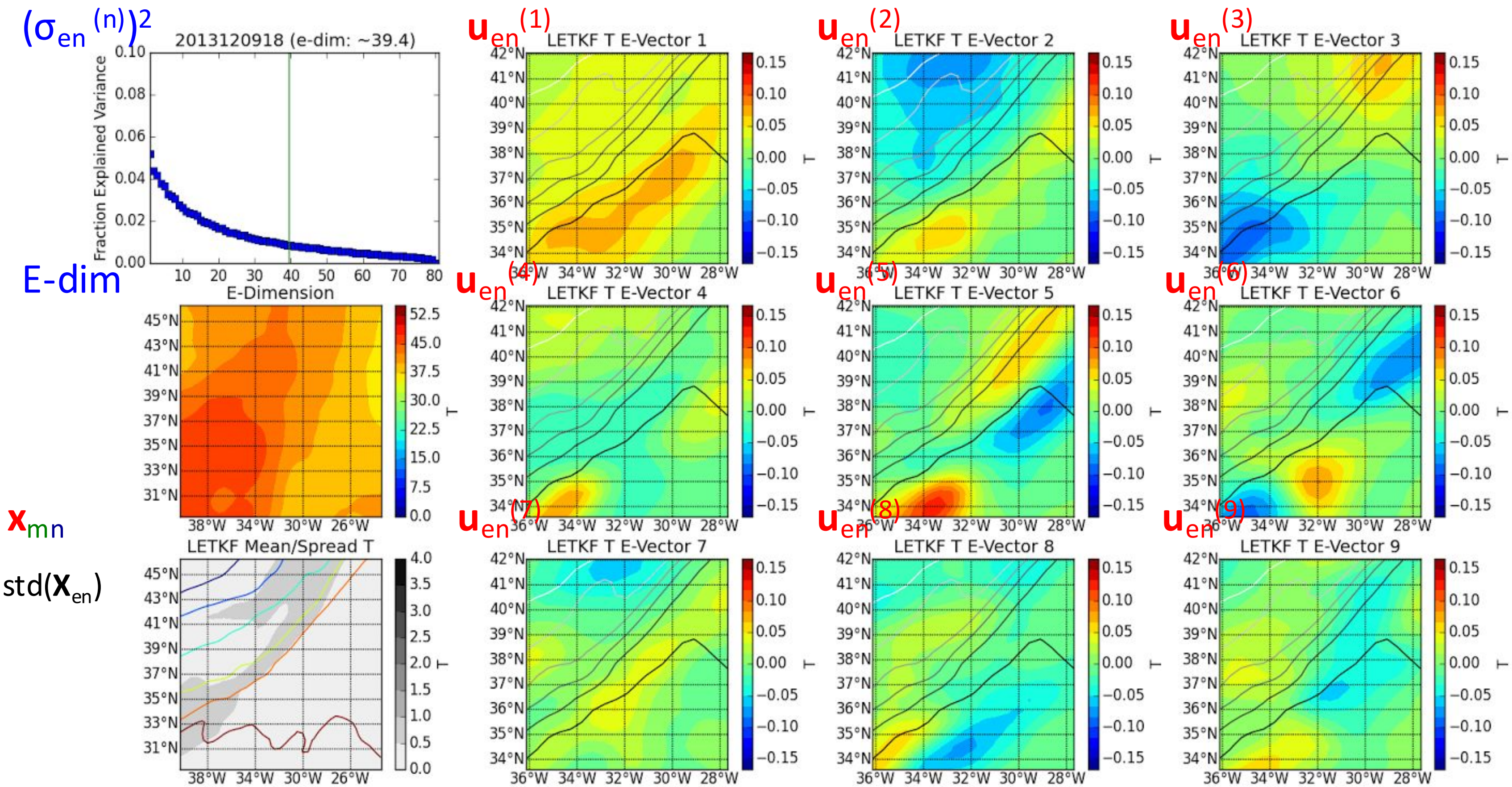
EnSRF vs LETKF: Background

- EnSRF: Local E-dim 34.3



EnSRF vs LETKF: Background

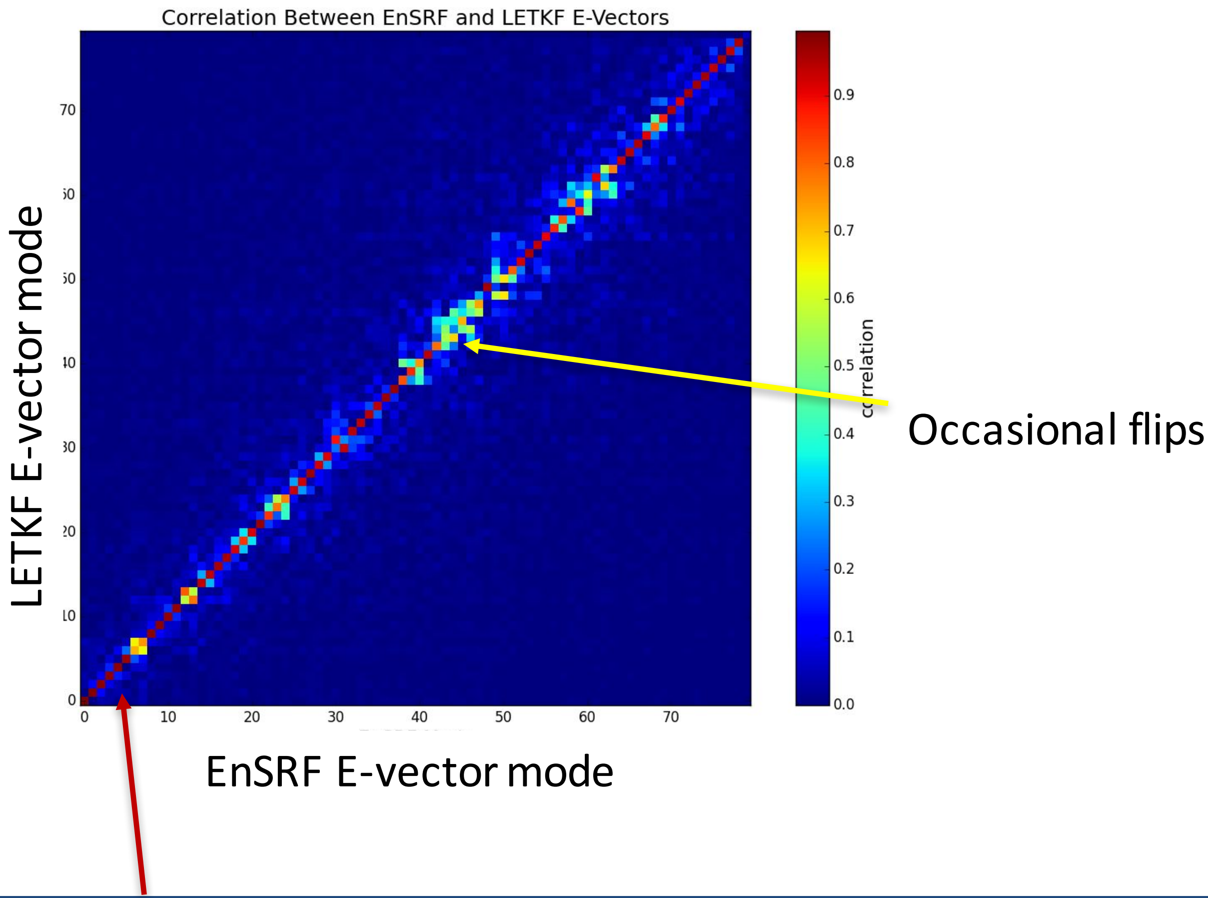
- LETKF: Local E-dim 39.4



Very similar - sign can flip

EnSRF vs LETKF: Background

- Local E-vector correlation

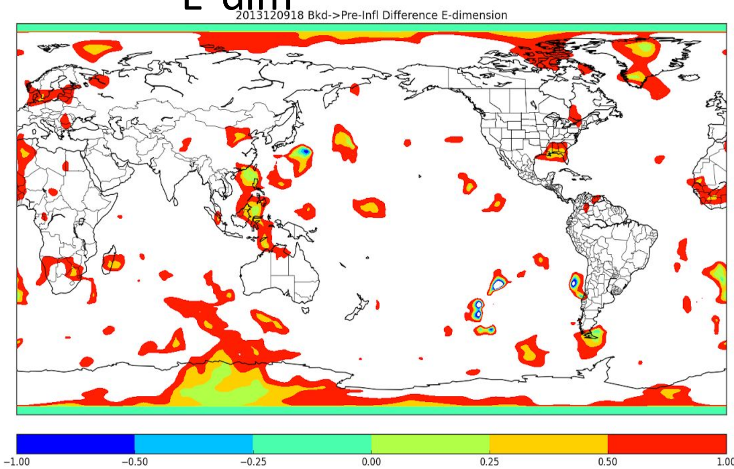


EnSRF and LETKF have very similar ensemble perturbation structures, from **P**-view point

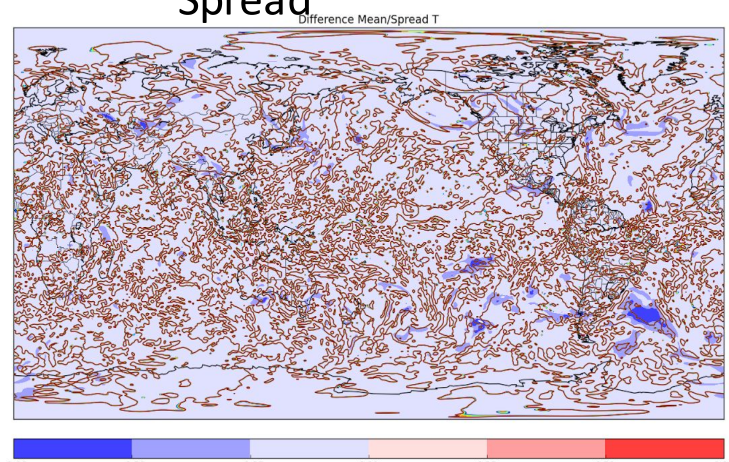
EnSRF Analysis

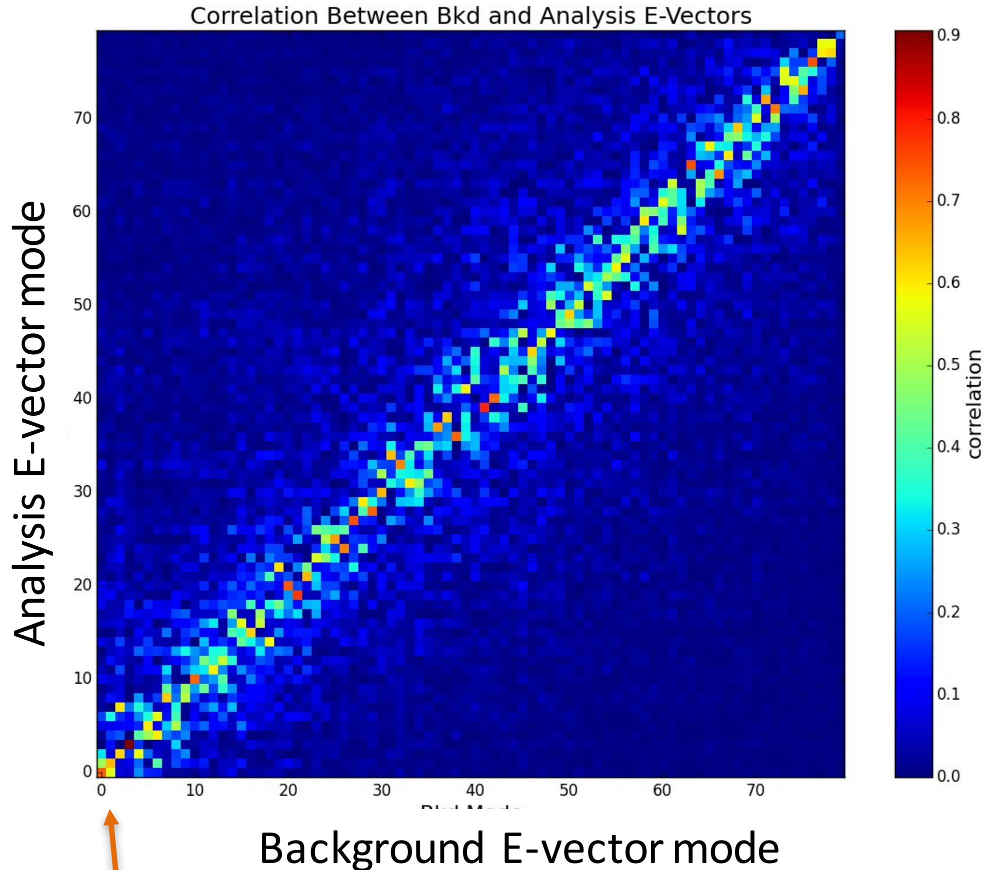
- Difference: Background-Analysis

E-dim



Spread



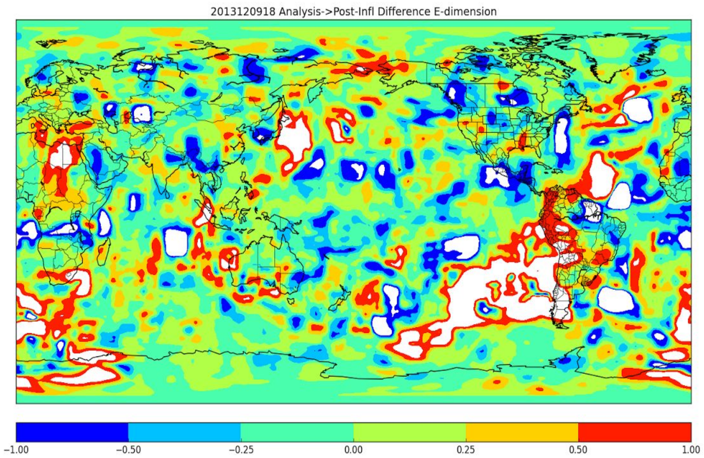


Knocking down “errors of the day”

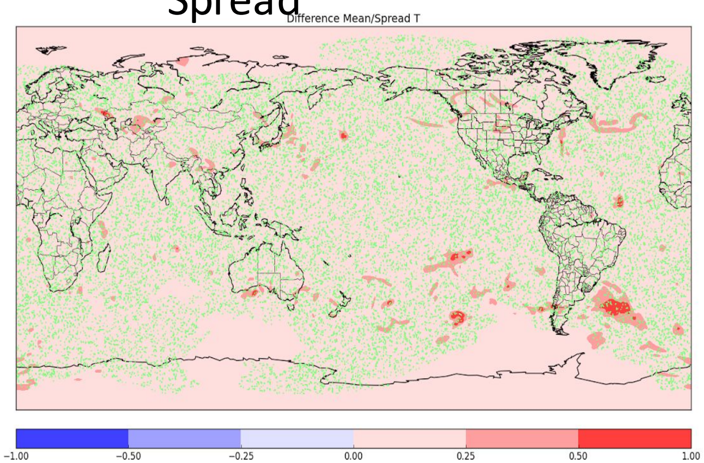
EnSRF Analysis

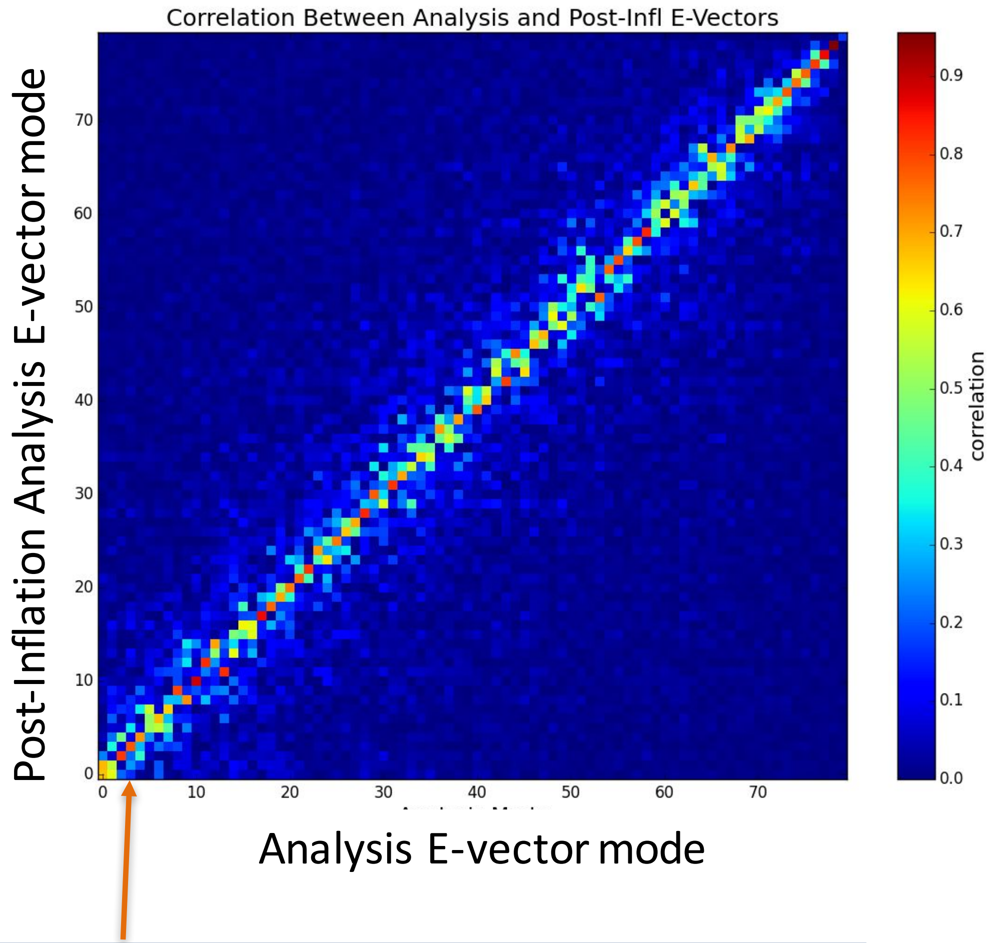
- Difference: Analysis pre- & post- inflation

E-dim



Spread





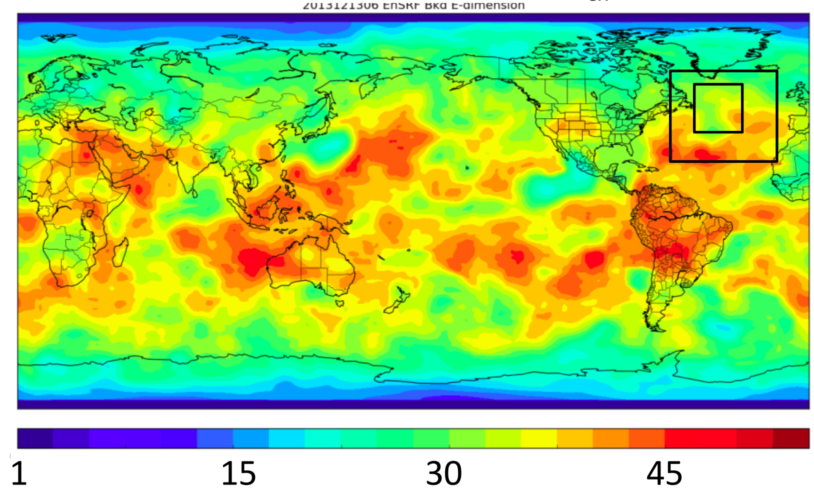
RTPS putting back "errors of the day"???

Localization Size

- 20131209 18z: T at 500 mb
  - $\mathbf{X}_{en}$: N x M dimensional
    - N=1083=19*19*3 (u,v,T)
    - M=80 ensemble
  - Background
    - Ensemble mean
    - Ensemble spread

Low-dim dynamic instability?

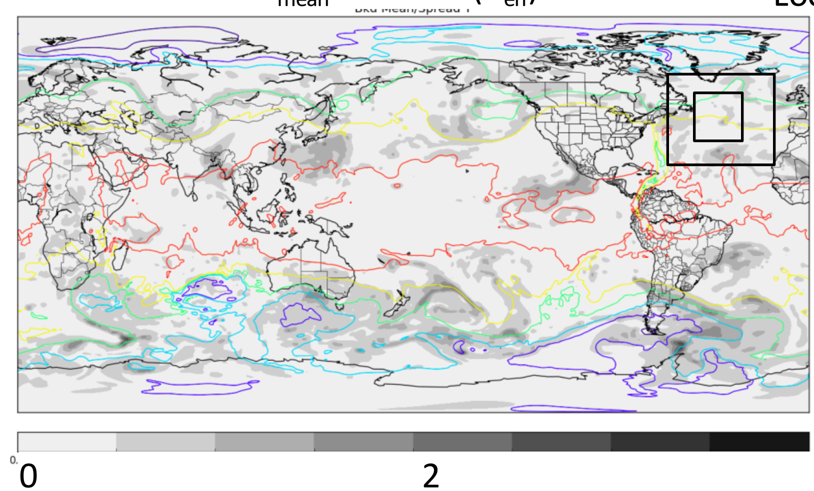
Local E-dim based on $\mathbf{X}_{en}$



Locally lower E-dim?

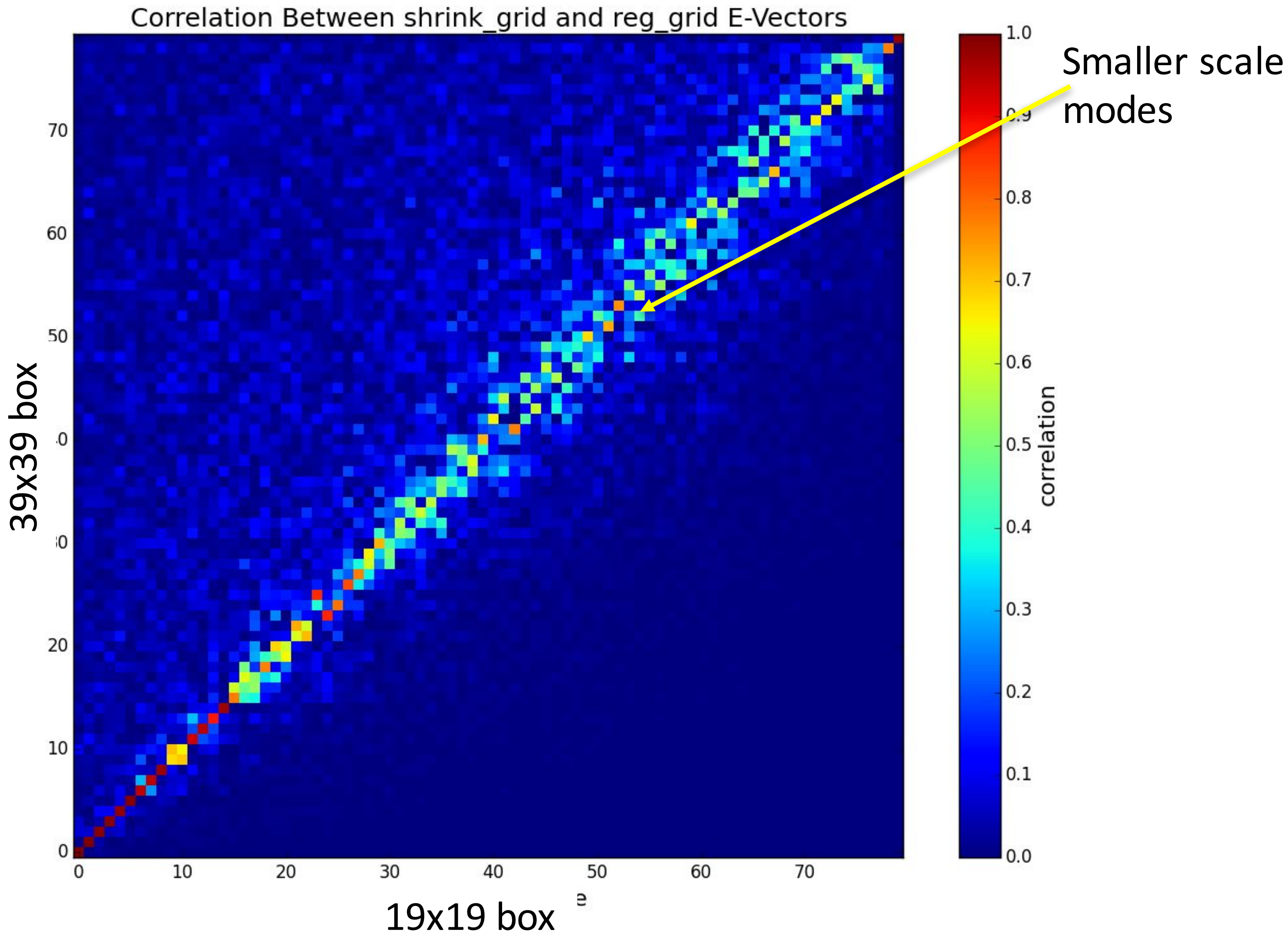
Locally large spread?

$\mathbf{x}_{mean}$ and var($\mathbf{X}_{en}$)



Large uncertainty

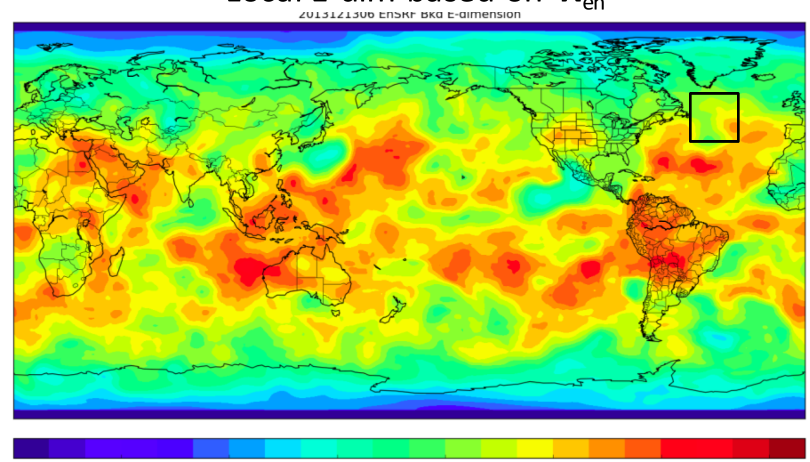
Localization Size



Local E-Dimension and Synoptic Conditions

- 20131209 18z: T at 500 mb
  - $\mathbf{X}_{en}$: N x M dimensional
    - N=1083=19*19*3 (u,v,T)
    - M=80 ensemble
  - Background
    - Ensemble mean
    - Ensemble spread

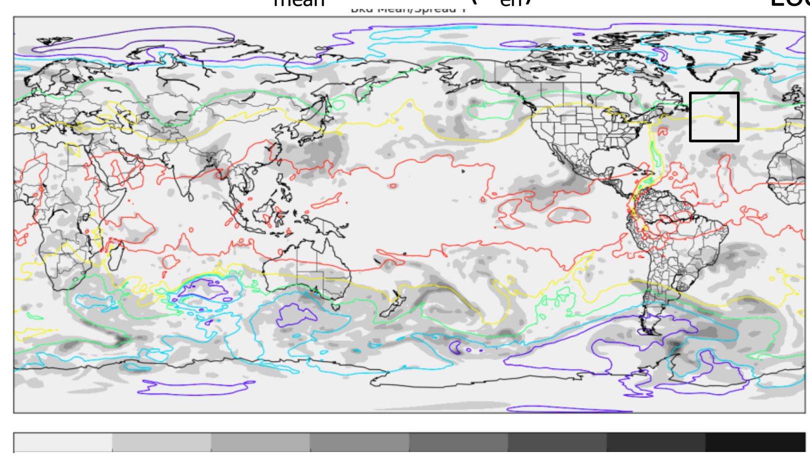
Local E-dim based on $\mathbf{X}_{en}$



Locally lower E-dim?

Low-dim dynamic instability?

$\mathbf{x}_{mean}$ and var($\mathbf{X}_{en}$)

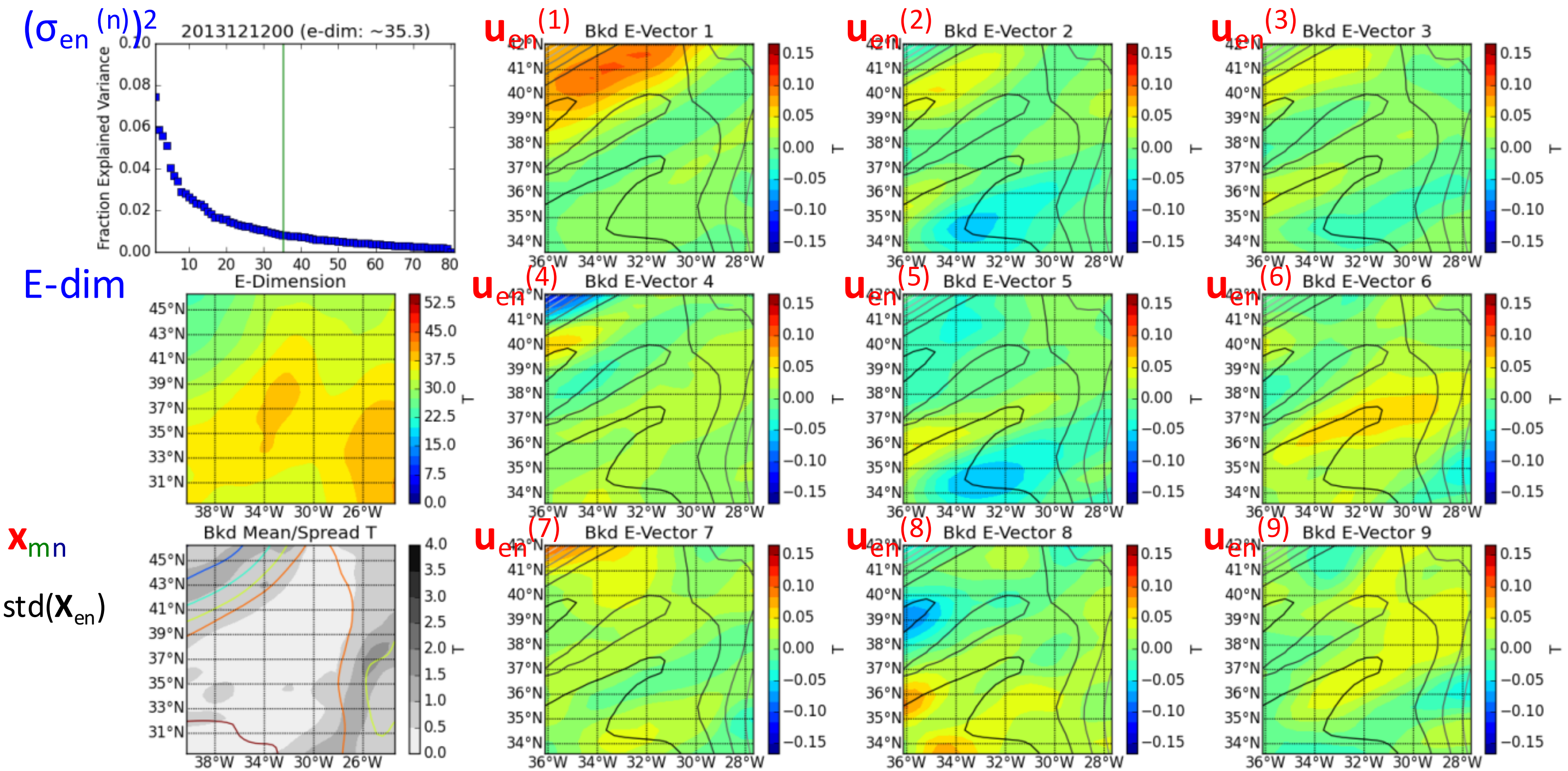


Locally large spread?

Large uncertainty

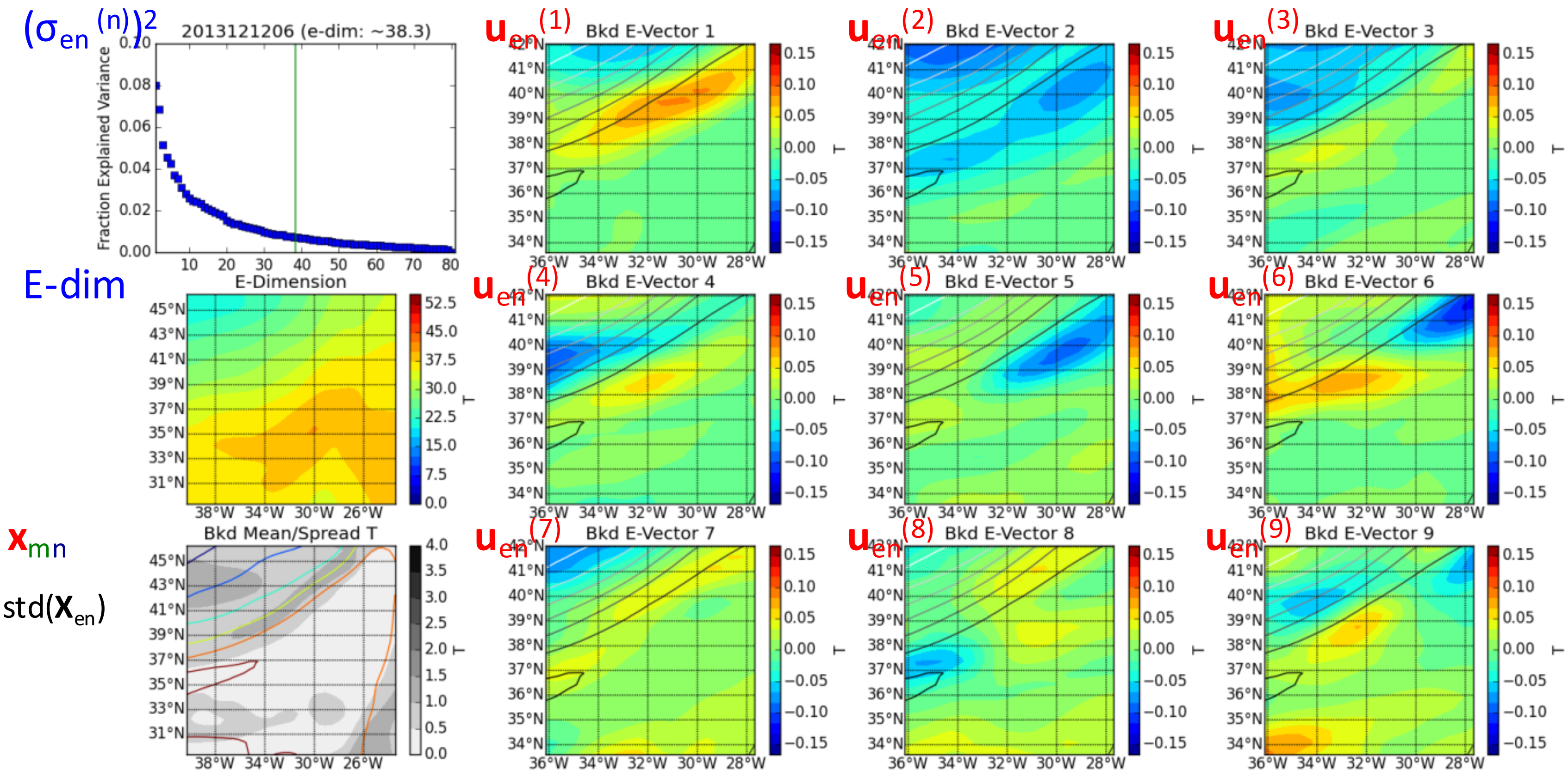
Case Study: Evolution 500mb *T* local E-diagnostics

- 2013121200: E-dim=35.3



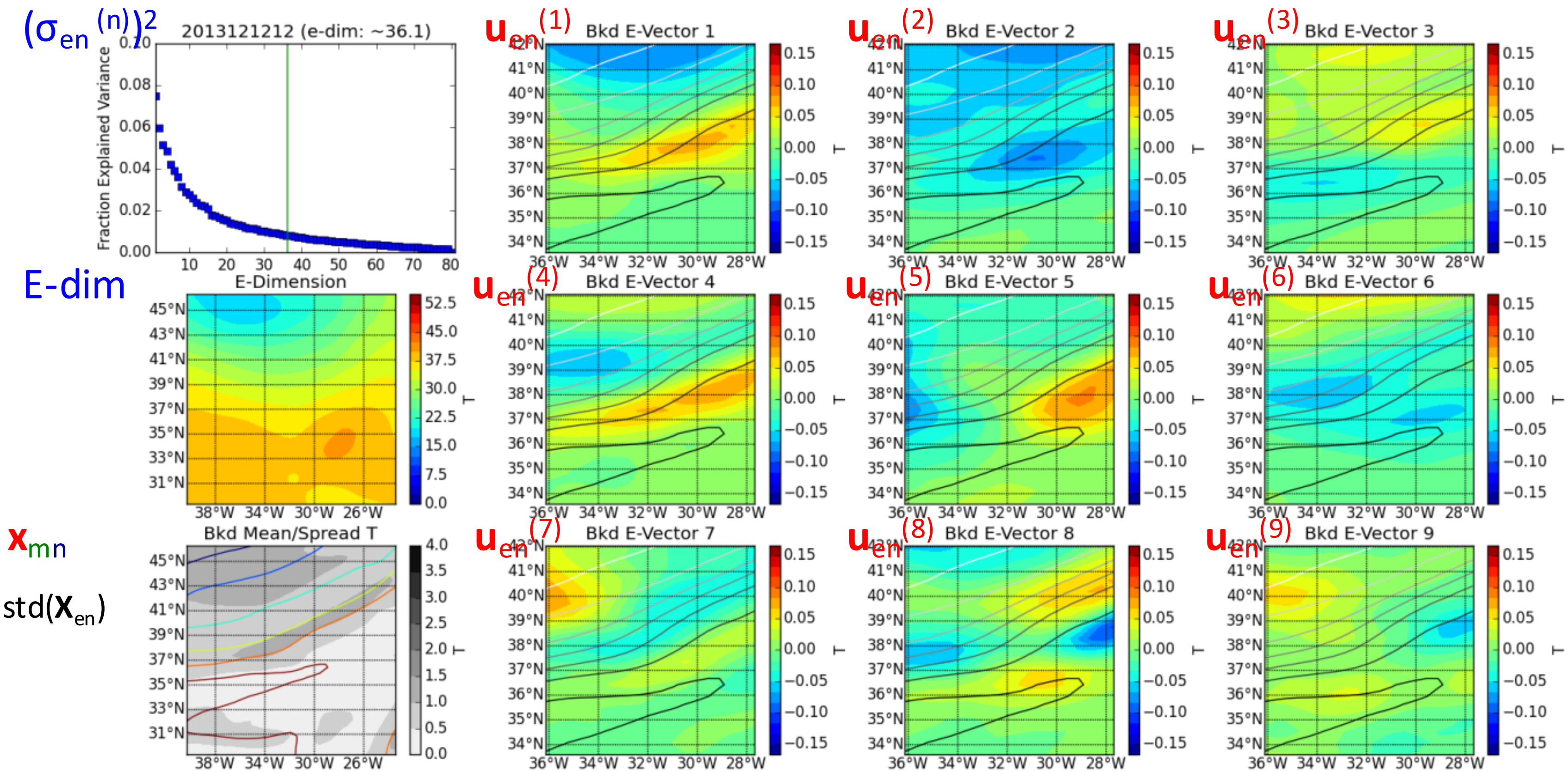
Case Study: Evolution 500mb T local E-diagnostics

- 2013121206: E-dim=36.8



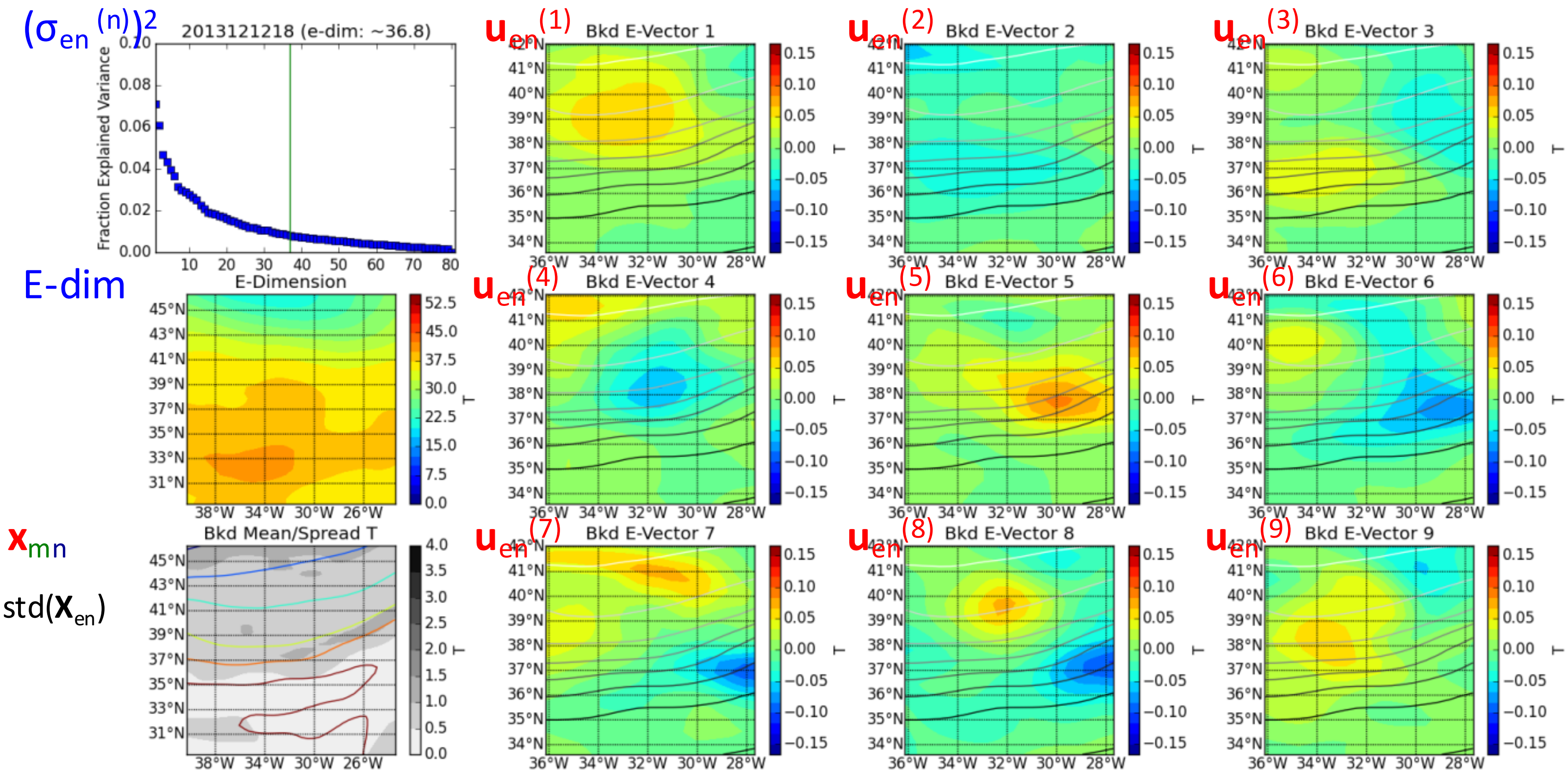
Evolution 500mb *T* local E-diagnostics

- 2013121212: E-dim=36.1



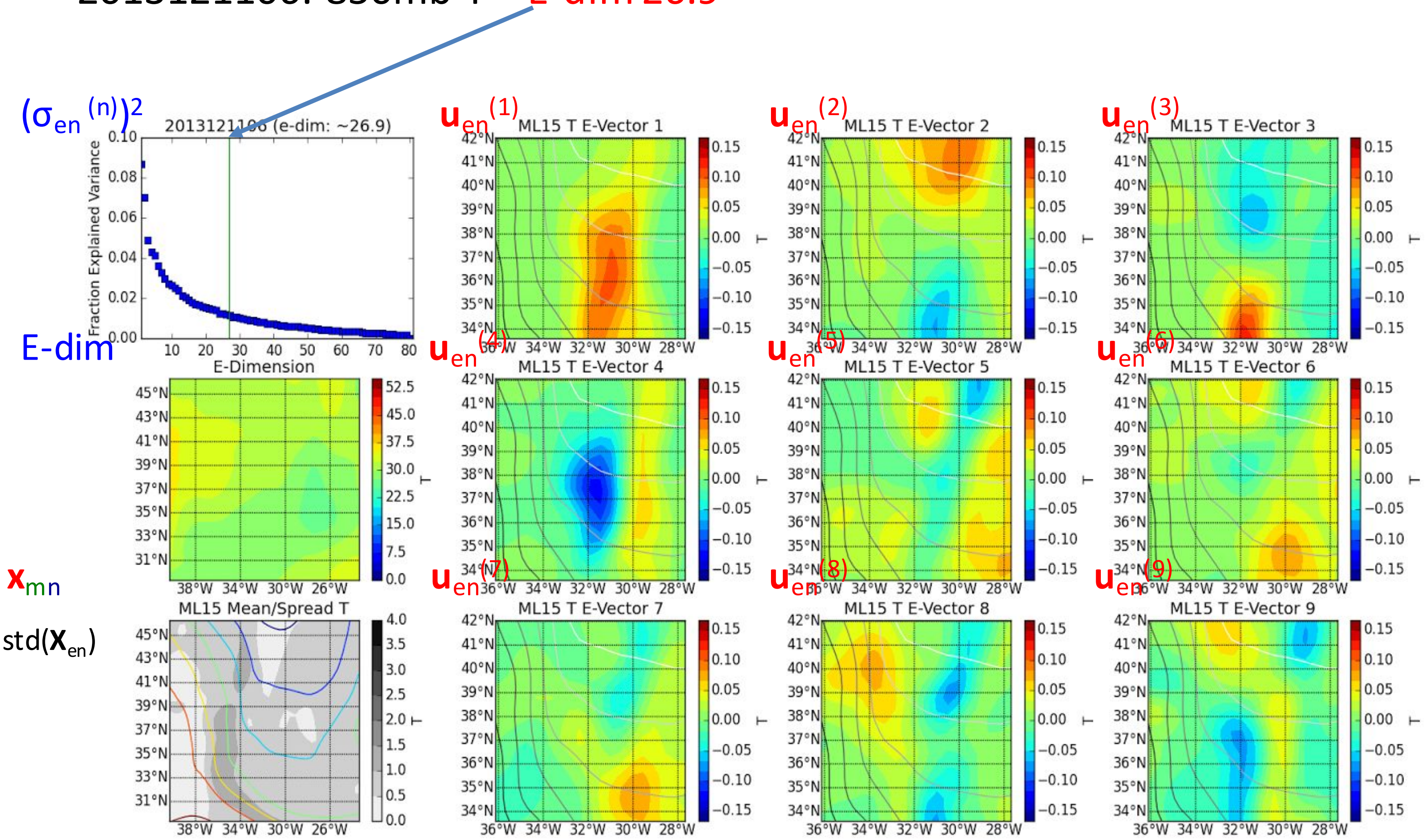
Evolution 500mb *T* local E-diagnostics

- 2013121218: E-dim=36.8



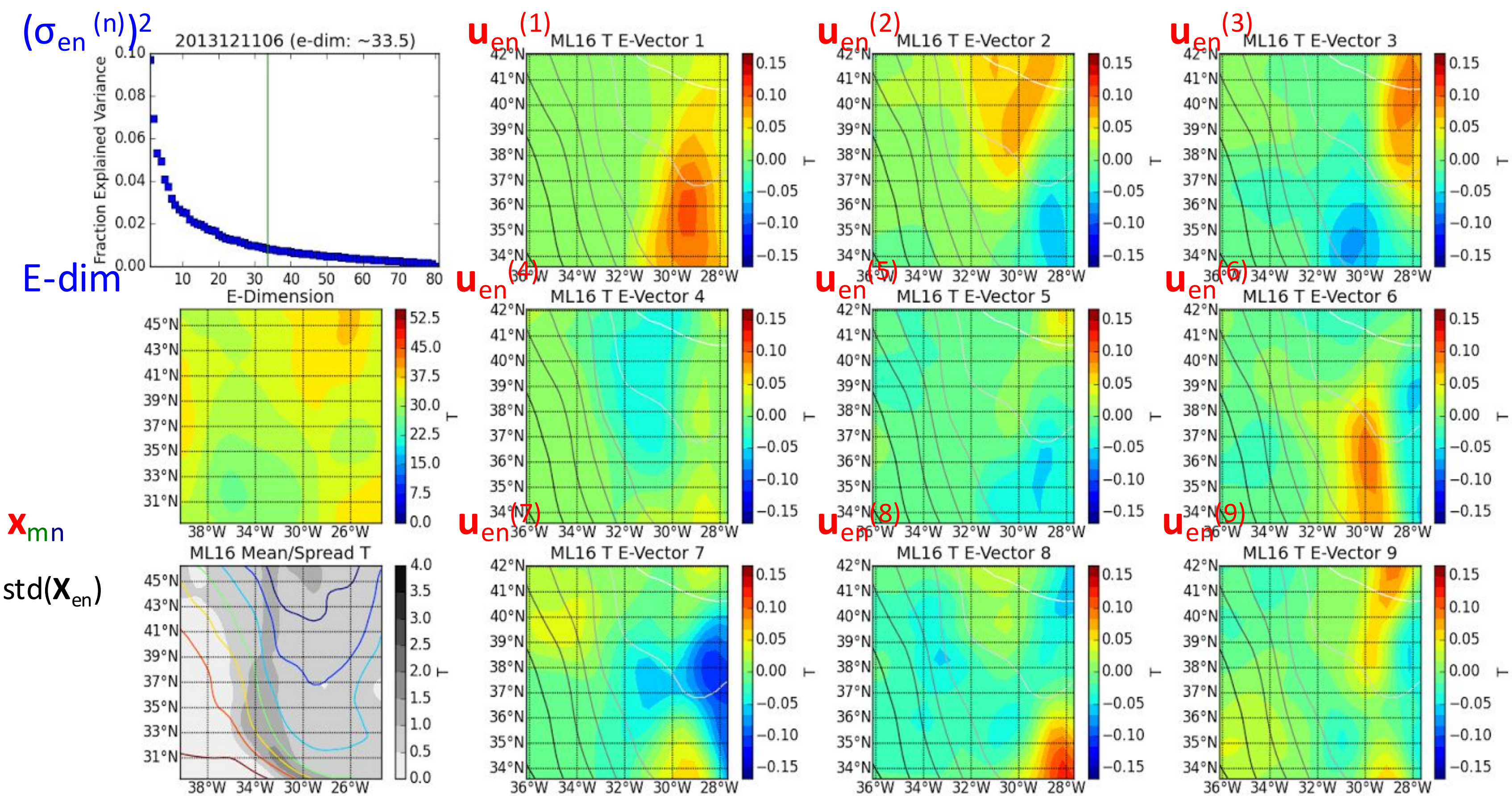
Vertical Variation

- 2013121106: 850mb T E-dim 26.9



Vertical Variation

- 2013121106: 825mb T   E-dim 33.5



Eastward tilt between 825 and 850mb

Summary

- **Local E-diagnostics:** One way to diagnose
  - Local instability represented by ensemble spread

- **Application to GSI**
  - Based on **P**, EnSRF and LETKF are extremely similar
    - Due to the same inflation (RTPS)?

  - Background vs Analysis
    - DA supresses the “erros of the day”.
    - Impact of stochastic perturbation in model forecast?

Thank You.

Questions?